AVALIAÇÃO DA MATÉRIA "INFERÊNCIA I"

Felipe Neres Silva Bezerra

Exercício 6: O nível de colesterol no sangue é uma variável aleatória com distribuição normal, de média desconhecida μ e desvio-padrão $\sigma = 60mg/100ml$.

a) Qual um estimador para a média μ?

O estimador mais assertivo para a média populacional μ é a média amostral $\bar{X}$.

b) Você conhece a distribuição desse estimador?

A distribuição da média amostral poderá ser considerada como uma Distribuição Normal se houver ao menos cerca de 30 observações na amostra.

c) Em uma amostra de 50 pacientes, observou-se uma média amostral $\bar{X} = 268$. Qual seria uma estimativa pontual para μ?

Estima-se que $\mu = 268$.

d) Considerando a amostra do item anterior, construa um intervalo de confiança para a média desconhecida com nível de confiança de 95 %.

$$\gamma = 0{,}95 = \mathbb{P}\left[\bar{x} - \frac{z\sigma}{\sqrt{n}} ; \bar{x} + \frac{z\sigma}{\sqrt{n}}\right] = \mathbb{P}\left[268 - \frac{1{,}96.60}{\sqrt{50}} ; \bar{x} + \frac{1{,}96.60}{\sqrt{50}}\right]$$
$$= \mathbb{P}[251{,}3688 ; 284{,}631]$$

e) Teste a hipótese de que $\mu = 260$, contra a alternativa de que $\mu > 260$ com base na mesma amostra. Utilize um nível de 5%.

$$\begin{cases} H_0 : \mu = 260 \\ H_1 : \mu > 260 \end{cases} \qquad RC = \{\bar{X} > x_c\}$$

$$\bar{X} \sim Normal(260{,}3600)$$

$$\alpha = \mathbb{P}[\bar{X} > x_c | \mu = 260] = \mathbb{P}\left[\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{x_c - \mu}{\sigma/\sqrt{n}} \middle| \mu = 260\right]$$

$$= \mathbb{P}\left[Z > \frac{x_c - 260}{60/\sqrt{50}} \middle| \mu = 260\right] = 0{,}05$$

$$1{,}64 = \frac{x_c - 260}{60/\sqrt{50}}$$

$$x_c = 260 + \frac{16{,}4.60}{\sqrt{50}} = 273{,}92 mg/100ml$$

$$RC = \{\bar{X} > 273{,}92 mg/100ml\}$$

Conforme a média amostral observada $\bar{X} = 268$ não pertence à região crítica, pode-se aceitar a hipótese nula de que a média do nível de colesterol no sangue equivale a $260 mg/100ml$

f) Qual deve ser o tamanho da amostra para que o intervalo de confiança tenha um erro de 15 unidades? Use 95 %de confiança.

$$Erro = \frac{Z.\sigma}{\sqrt{n}}$$

$$n = \left(\frac{Z.\sigma}{Erro}\right)^2 = \left(\frac{1{,}96 \, . \, 60}{15}\right)^2 = 61{,}4656$$

Para um erro de estimativa de, no máximo, 15 unidades, necessita-se de uma amostra com, ao menos, 62 observações.

Exercício 8: Suponha que se deseje estimar a proporção p de indivíduos com certa doença em uma dada região. Selecionou-se uma amostra aleatória de 100 pessoas e constatou-se que 25 eram portadoras de doença.

a) Calcule a estimativa pontual da proporção p.

$$\hat{p} = \frac{X}{n} = \frac{25}{100} = 0{,}25$$

b) Construa o intervalo de confiança para p com coeficiente de confiança de 95%. Qual o comprimento do intervalo?

$$\gamma = \mathbb{P}\left[\hat{p} - z.\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + z.\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$
$$= \mathbb{P}\left[0,25 - 1,96.\sqrt{\frac{0,25(1-0,25)}{100}}; 0,25 + 1,96.\sqrt{\frac{0,25(1-0,25)}{100}}\right]$$
$$= \mathbb{P}[0,1651; 0,3349]$$

c) Um pesquisador acredita que a proporção de doentes é superior a 20%. Teste essa hipótese ao nível de $\alpha = 0,05$. Formule as hipóteses nula e alternativa.

$$\begin{cases} H_0: p = 0,2 \\ H_1: p > 0,2 \end{cases} \qquad RC = \{\hat{p} > p_c\}$$
$$\hat{p} \sim Normal\left(0,2; \frac{0,2.0,8}{100}\right)$$

$$\alpha = \mathbb{P}[\hat{p} > p_c | p = 0,2] = \mathbb{P}\left[\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} > \frac{x_c - \mu}{\sqrt{p(1-p)/n}} \middle| p = 0,2\right]$$
$$= \mathbb{P}\left[\left\{Z > \frac{p_c - 0,2}{\sqrt{0,2.0,8/100}} \middle| p = 0,2\right\}\right] = 0,05$$

$$1,64 = \frac{p_c - 0,2}{\sqrt{0,2.0,8/100}}$$
$$p_c = 0,2 + 1,64\sqrt{\frac{0,2.0,8}{100}} = 0,2656$$

$$RC = \{\hat{p} > 0,2656\}$$

Conforme observado na amostra $\hat{p} = 0,25 < 0,2656$, este valor não encontra-se na região crítica. Com isso, aceita-se H_0 e conclui-se que a proporção de doentes não é superior a 20%.

Exercício 11: O número de pontos em um exame de inglês tem sido historicamente ao redor de 80. Sorteamos 10 estudantes que fizeram recentemente esse exame e observamos as notas: 65; 74; 78; 86; 59; 84; 75; 72; 81, e 83.

a) Qual uma estimativa pontual para a nota média?
A estimativa mais assertiva para a média populacional μ é a dada pelo estimador média amostral $\bar{X}$.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{65 + 74 + 78 + 86 + 59 + 84 + 75 + 72 + 81 + 83}{10} = 75,7$$

b) Construa um intervalo de confiança para a nota média com 95% de confiança.

Estima-se a variância através de S^2:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})$$

$$S^2 = ((65 - 75,7)^2 + (74 - 75,7)^2 + (78 - 75,7)^2 + (86 - 75,7)^2 + (59 - 75,7)^2 + (84 - 75,7)^2 + (75 - 75,7)^2 + (72 - 75,7)^2 + (81 - 75,7)^2 + (83 - 75,7)^2)/9 = 74,6778$$

c) Especialistas desconfiam que a média diminuiu e desejam testar essa afirmação através de um teste de hipóteses, com nível de significância de 5%. Fazendo as suposições necessárias, qual seria a conclusão do teste?

$$\begin{cases} H_0: \mu = 80 \\ H_1: \mu < 80 \end{cases} \qquad RC = \{T < t_c\} = \{T < -1,833\}$$

$$T = \frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim t_9$$

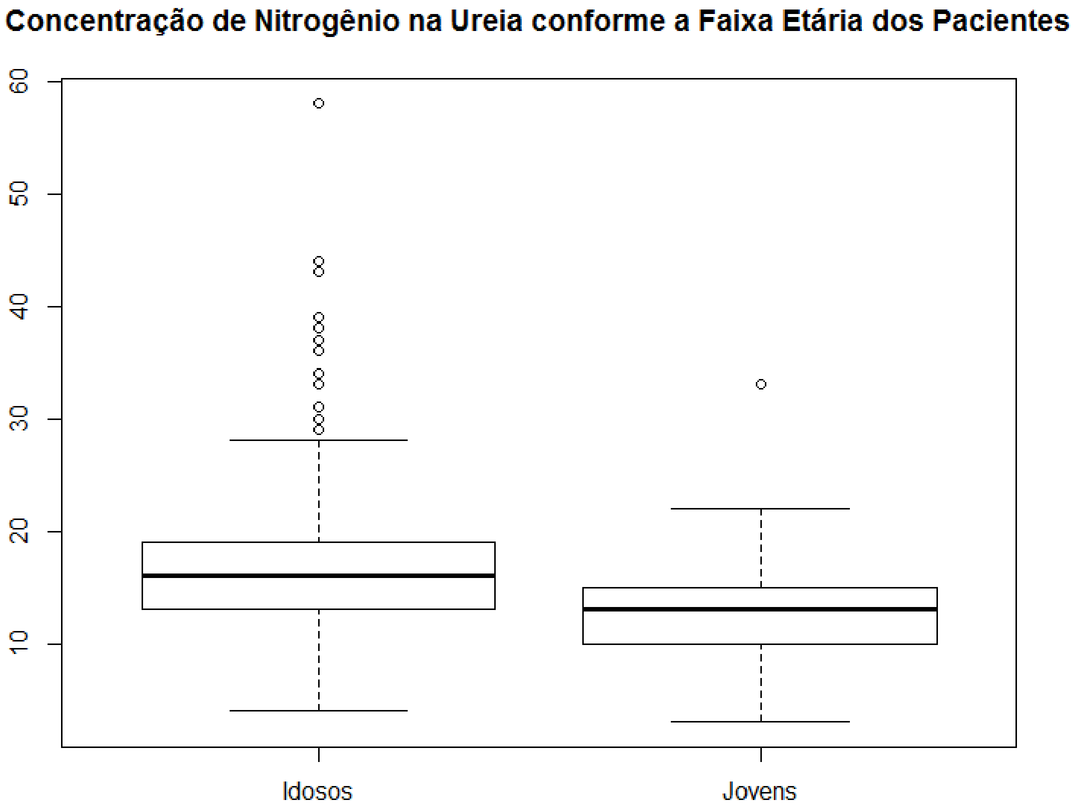
$$T = \frac{75,7 - 80}{\sqrt{74,6778/9}}$$

$$T = -1,4928$$

Como $T = -1,4928$ não pertence à região crítica, H_0 é aceita e conclui-se que a média do número de pontos do exame de inglês não diminuiu.

Exercício 13: Considerando os dados do arquivo *cancer.xlsx*, defina dois grupos de pacientes: um de jovens, com idade inferiores ou iguais a 54 anos, e um de idosos, com idades superiores a 54 anos. Os grupos deverão conter 191 e 171 pacientes. Considere a variável nitrogênio na ureia (N).

a) Construa um box-plot para a variável N para cada um dos grupos etários e compare-os descritivamente. Com base nos gráficos, existem indicações de que a idade está influenciando a concentração de nitrogênio na ureia?



Conforme o gráfico box-plot sugere, há indícios de que pacientes com idade superior a 54 anos tenham maior concentração de nitrogênio na ureia, tendo em vista as medidas de mediana, primeiro quartil e terceiro quartil encontrarem-se mais elevadas que na amostra de pacientes mais jovens.

b) É de interesse verificar se a média populacional da variável N para os pacientes idosos é superior a 15. Sendo a variável desconhecida, qual conclusão pode ser obtida para um nível de significância de 5%?

t = 3.9029, df = 170, p-value = 6.841e-05
alternative hypothesis: true mean is greater than 15
95 percent confidence interval:
 16.28391   Inf
sample estimates:
mean of x
 17.22807

$$\begin{cases} H_0: \mu = 15 \\ H_1: \mu < 15 \end{cases}$$

$$\alpha^* = \mathbb{P}[\bar{X} > 17{,}22807 | \mu = 15] = 0{,}00006841$$

Considerando que a probabilidade de que a média amostral $\bar{X} > 17{,}22807$ enquanto a média populacional $\mu = 15$ é de apenas $0{,}006841$, estabelecendo o nível de significância em 5%, rejeita-se a hipótese nula e admite-se que a média populacional da variável nitrogênio na ureia (N) para os pacientes idosos é superior a 15.

c) Considerando agora o grupo de pacientes mais jovens, verifique se a média populacional para N é menor do que 15. Obtenha o nível descritivo e conclua ao nível de 5%.

t = -7.7181, df = 190, p-value = 3.263e-13
alternative hypothesis: true mean is less than 15
95 percent confidence interval:
-Inf 13.28844
sample estimates:
mean of x
12.82199

$$\begin{cases} H_0: \mu = 15 \\ H_1: \mu < 15 \end{cases}$$

$$\alpha^* = \mathbb{P}[\bar{X} < 12{,}82199 | \mu = 15] = 3{,}263.10^{-13}$$

Considerando que a probabilidade de que a média amostral $\bar{X} < 12{,}82199$ enquanto a média populacional $\mu = 15$ é de apenas $3{,}263.10^{-11}$%, estabelecendo o nível de significância em 5%, rejeita-se a hipótese nula de que e admite-se que a média populacional da variável nitrogênio na ureia (N) para os pacientes jovens é inferior a 15.

d) Construa intervalos de confiança para a média populacional da variável N para os dois grupos com 95% de confiança. Compare os intervalos.

Intervalo de confiança para a média populacional da variável N para o grupo de pacientes idosos: 16,10114 a 18,35500.

Intervalo de confiança para a média populacional da variável N para o grupo de pacientes jovens: 12,26535 a 13,37863.

e) Com base nos resultados dos itens B e C, discuta o comportamento das médias da variável N para os dois grupos de pacientes.

Em concordância com os resultados obtidos nos itens b) e c), e conforme já concluído nestes, constata-se que a média populacional da variável “nitrogênio na ureia” (N) para os pacientes jovens encontra-se abaixo de

15, enquanto que a média populacional da mesma variável para pacientes idosos deve situar-se acima deste patamar.

Exercício 14: Um criador tem constatado uma porção de 10% do rebanho com verminose. O veterinário alterou a dieta dos animais e acredita que a doença diminuiu de intensidade. Um exame de 100 cabeças do rebanho, escolhidas ao acaso, indicou 8 delas com verminose. Ao nível de 5%, há indícios de que a proporção diminuiu?

```
data:  8 out of 100, null probability 0.1
X-squared = 0.44444, df = 1, p-value = 0.2525
alternative hypothesis: true p is less than 0.1
95 percent confidence interval:
 0.0000000 0.1364648
sample estimates:
   p
0.08
```

$$\begin{cases} H_0: p = 0{,}1 \\ H_1: p < 0{,}1 \end{cases}$$

$$\alpha^* = \mathbb{P}[\bar{p} < 0{,}08 | p = 0{,}1] = 0{,}2525$$

Considerando que a probabilidade de que a proporção amostral $\bar{p} < 0{,}08$ enquanto a proporção populacional $p = 0{,}1$ é de $25{,}25\%$, estabelecendo o nível de significância em 5%, aceita-se a hipótese nula e admite-se que a proporção populacional de cabeças de gado com verminose não diminuiu.